



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/762,126	02/02/2001	Shinji Miwa	P5276b	1392

20178 7590 05/07/2004

EPSON RESEARCH AND DEVELOPMENT INC  
INTELLECTUAL PROPERTY DEPT  
150 RIVER OAKS PARKWAY, SUITE 225  
SAN JOSE, CA 95134

EXAMINER

BLACKWELL, JAMES H

ART UNIT	PAPER NUMBER
----------	--------------

2176

DATE MAILED: 05/07/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

## Office Action Summary

Application No.

09/762,126

Applicant(s)

MIWA ET AL.

Examiner

James H Blackwell

Art Unit

2176

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

### Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

### Status

- 1) ☒ Responsive to communication(s) filed on 02 February 2001.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

### Disposition of Claims

- 4) ☒ Claim(s) 1-13 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 1-4 and 6-13 is/are rejected.
- 7) ☒ Claim(s) 5 is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

### Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☒ The drawing(s) filed on 02 February 2001 is/are: a) ☒ accepted or b) ☐ objected to by the Examiner.  
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

### Priority under 35 U.S.C. § 119

- 12) ☒ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some \* c) ☐ None of:
- ☐ Certified copies of the priority documents have been received.
  - ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
  - ☒ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

### Attachment(s)

- 1) ☒ Notice of References Cited (PTO-892)
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3) ☒ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)  
Paper No(s)/Mail Date \_\_\_\_\_
- 4) ☐ Interview Summary (PTO-413)  
Paper No(s)/Mail Date. \_\_\_\_\_
- 5) ☐ Notice of Informal Patent Application (PTO-152)
- 6) ☐ Other: \_\_\_\_\_

## DETAILED ACTION

### *Claim Objections*

Claim 5 is objected to under 37 CFR 1.75(c) as being in improper form because multiple dependent claim cannot depend from any other multiple dependent claim. See **MPEP § 608.01(n)**. Accordingly, Claim 5 has not been further treated on the merits.

### *Claim Rejections - 35 USC § 102*

1. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(a) the invention was known or used by others in this country, or patented or described in a printed publication in this or a foreign country, before the invention thereof by the applicant for a patent.

Claims 1-4, 6, and 10-13 are rejected under 35 U.S.C. 102(a) as being anticipated by Zamir et al. (hereinafter Zamir, "Web Document Clustering: A Feasibility Demonstration", ACM, August, 1998).

In regard to independent Claim 1, Zamir teaches the STC algorithm which is a linear time clustering algorithm. STC has three logical steps: (1) document cleaning, (2) identifying base clusters using a suffix tree, and (3) merging the base clusters into clusters (p. 48, Col. 1, Sec. 3, lines 18-25; compare to Claim 1, "***A document categorizing method for categorizing a plurality of documents into a plurality of clusters according to semantic similarity, and said method being characterized in that: ...***"). Zamir also teaches that step (2) of the STC algorithm, the identification of base clusters can be viewed as the creation of an inverted index of phrases for our

document collection. This is done efficiently using a data structure called a *suffix tree*. This structure can be constructed in time linear with the size of the collection, and can be constructed incrementally as the documents are being read (p. 48, Col. 1, Sec 3.2, lines 43-49). Each base cluster is assigned a score that is a function of the number of documents it contains, and the number of words that make up its phrase (p. 48, Col. 2, Sec 3.2, lines 30-32; compare to Claim 1, “... ***after categorizing said plurality of documents into a plurality of clusters according to semantic similarity, a cluster merging process is performed such that relations among clusters of said plurality of clusters are evaluated on the basis of documents included in the respective clusters, ...***”). Zamir also teaches that the final step of the STC algorithm merges base clusters with a high degree of overlap in their document sets (p. 49, Col. 1, lines 19-21; compare to Claim 1, “... ***and two or more clusters having a degree of relation equal to or higher than a predetermined value are combined together***”).

In regard to dependent Claim 2, Zamir teaches that a binary similarity measure between base clusters based on the overlap of their document sets can be defined. Given two base clusters  $B_m$  and  $B_n$ , with sizes  $|B_m|$  and  $|B_n|$  respectively, and  $|B_m \text{ union } B_n|$  representing the number of documents common to both base clusters, we define the similarity of  $B_m$  and  $B_n$  to be 1 iff  $|B_m \text{ union } B_n| / |B_m| > 0.5$  and  $|B_m \text{ union } B_n| / |B_n| > 0.5$ . Otherwise, their similarity is defined to be zero (p. 49, Col. 1, Sec. 3.3, lines 24-33; compare to Claim 2, “... ***said cluster merging process is performed such that the evaluation of relations among clusters under consideration as to whether they should be merged or not is performed on the basis of the number of***”).

***documents commonly included in said clusters under consideration relative to the total number of documents included in said clusters under consideration, and cluster merging is performed in accordance with the evaluation result").***

In regard to dependent Claim 3, Zamir teaches that each base cluster is assigned a score that is a function of the number of documents it contains, and the number of words that make up its phrase (p. 48, Col. 2, Sec 3.2, lines 30-32; compare to Claim 3, ***"... said cluster merging process is performed such that in what manner feature elements, which characterize respective clusters under consideration as to whether they should be merged or not, appear in the respective clusters under consideration is examined, and cluster merging is performed in accordance with the manner in which the feature elements appear")***).

In regard to dependent Claim 4, Zamir teaches that in essence, we are clustering the base clusters using the equivalent of a single-link clustering algorithm where a predetermined minimal similarity between base clusters serves as the halting criterion (implying that it keeps clustering clusters until a condition is met) (p. 49, Col. 1, Sec 3.3, lines 40-41; Col. 2, lines 1-2; compare with Claim 4, ***"A document categorizing method according to one of Claims 1 to 3, wherein said cluster merging process is performed at least for two clusters, and after completion of the cluster merging process a first time, the cluster merging process is performed repeatedly for the resultant set of clusters until no further cluster merging occurs")***).

In regard to independent Claim 6 (and similarly independent Claims 11, and 13), Zamir teaches the STC algorithm which is a linear time clustering algorithm. STC has

three logical steps: (1) document cleaning, (2) identifying base clusters using a suffix tree, and (3) merging the base clusters into clusters (p. 48, Col. 1, Sec. 3, lines 18-25; compare to Claim 6 (and similarly Claims 11, and 13), ***“A document categorizing method for categorizing a plurality of documents into a plurality of clusters according to semantic similarity, said method being characterized in that: ...”***).

Zamir also teaches that step (2) of the STC algorithm, the identification of base clusters can be viewed as the creation of an inverted index of phrases for our document collection. This is done efficiently using a data structure called a *suffix tree*. This structure can be constructed in time linear with the size of the collection, and can be constructed incrementally as the documents are being read (p. 48, Col. 1, Sec 3.2, lines 43-49). Each base cluster is assigned a score that is a function of the number of documents it contains, and the number of words that make up its phrase (p. 48, Col. 2, Sec 3.2, lines 30-32; compare to Claim 6 (and similarly Claims 11, and 13), ***“... after categorizing said plurality of documents into a plurality of clusters according to semantic similarity, a cluster merging process is performed such that relations among clusters of said plurality of clusters are evaluated on the basis of documents included in the respective clusters, ...”***). Zamir also teaches that the final step of the STC algorithm merges base clusters with a high degree of overlap in their document sets (p. 49, Col. 1, lines 19-21; compare to Claim 6 (and similarly Claims 11, and 13), ***“... and two or more clusters having a degree of relation equal to or higher than a predetermined value are combined together ...”***). Zamir also teaches Figure 1, which depicts the output of the MetaCrawler-STC clustering engine for the

query "salsa". In the figure, only the first five clusters are shown. The words in bold are the shared phrases found in the clusters. Note the descriptive power of phrases such as "Puerto Rico", "Latin Music", and "York Salsa Dancers". (p. 47, Col. 1, Fig. 2 and caption; compare to Claim 6 (and similarly Claims 11, and 13), **"... information representing which clusters have been merged together and also representing the degrees of relation among the merged clusters is generated and said information is output together with the categorization result to be presented to a user so that when final clusters obtained as a result of said cluster merging process are displayed, the user can see in what manner said cluster merging process has been performed to obtain said final cluster"**").

### **Claim Rejections - 35 USC § 103**

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

Claims 7-9 are rejected under 35 U.S.C. 103(a) as being unpatentable over Zamir.

In regard to dependent Claims 7-9, Zamir fails to specifically teach the various means to display the results of the document categorizing method based on their degree of similarity and whether they are displayed in an AND or and OR form, or how

brackets are used to distinguish the AND and OR forms. However, it would have been obvious to one of ordinary skill in the art at the time of invention to provide displays based on the relationships between clusters, in view of Zamir's disclosure, providing the benefit of having simplified the understanding of the search results for the user.



**Conclusion**

Any inquiry concerning this communication or earlier communications from the examiner should be directed to James H Blackwell whose telephone number is 703-305-0940. The examiner can normally be reached on Mon-Fri.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Joseph H Feild can be reached on 703-305-9792. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

James H. Blackwell  
04/21/04

  
**JOSEPH FEILD**  
**SUPERVISORY PATENT EXAMINER**